



Virident tachION Drive

User's Guide

© 2010 Virident Systems, Revision 1.4a

Table of Contents

Introduction	1
Specifications	2
System Requirements	2
Hardware.....	2
Operating System Support.....	2
Installation	3
Hardware installation.....	3
Software installation	4
Verify the Virident tachION is installed and detected properly	5
Choosing and installing the proper driver RPMs	5
SUSE Only: Enable unsupported modules to allow driver loading	6
RECOMMENDED: Enhance performance by disabling CPU throttling.....	6
Update kernel options and reboot	7
BIOS setup for optimal performance	7
Verify the device driver is loaded and functioning	8
Verify installation performance	9
Use the “/dev/vgca0” device	9
Partitioning and Configuring with VGC-Config	10
Partitioning and selecting write performance of the device	10
Usable space vs. write performance tradeoffs	10
Performance implications	10
VGC-Config example	11
Creating Filesystems	13
EXT3 high performance options.....	13
XFS high performance options	13

Software RAID	14
RAID-1	14
RAID-5/6.....	14
Multiple Virident tachION Cards	15
TachION Management Utilities	16
VGC-Config	16
VGC-Config card settings	17
VGC-Config partition settings	17
VGC-Diags.....	18
Performance Optimization.....	19
Disabling power saving and enabling performance CPU modes	19
Increasing thread counts to boost performance	19
MySQL specific optimization	20
Filesystem selection	20
Troubleshooting.....	21
Cannot find block device /dev/vgca0.....	21
Performance is lower than expected	22
IRQBalance not running	22
Appendix A: Onboard diagnostic LEDs.....	23

VIRIDENT SOFTWARE END USER LICENCE AGREEMENT

This is a legal agreement between you, the end user, ("You") and Virident Systems Inc. ("Virident"). BY INSTALLING THIS SOFTWARE OR CLICKING ON THE "ACCEPT" BUTTON YOU ARE AGREEING TO BE BOUND BY THE TERMS OF THIS AGREEMENT. Certain of the software delivered may be open source software and is licensed pursuant to the applicable open source licenses specified in the documentation. This Agreement applies to the software that is proprietary to Virident (the "Program(s)").

1. License

Virident grants You a non-exclusive, non-transferable license to use one copy of the Program(s) for each server product for which you have paid the applicable license fee, for internal business purposes, in accordance with the user documentation, and in accordance with the terms of this Agreement. You may make one back-up copy of the Program, provided that such copy shall be subject to the terms of this Agreement. You may not use the Program(s) with any equipment other than the equipment specified by Virident. You may not loan, rent, lease, license, sublicense or distribute the Program(s) or any copies.

3. Copyright and other intellectual property rights

All intellectual property rights in the Program(s) and user documentation are owned by Virident and are protected by copyright and other intellectual property laws, international treaty provisions and applicable national laws. All rights not expressly granted in this Agreement are reserved by Virident.

4. Further restrictions

You must not reverse engineer or decompile the Program(s) except to the extent permitted by applicable law despite any contractual prohibition. If you wish to obtain interface information relating to the Program, to achieve interoperability, a request should be made in writing to Virident.

5. No warranties

ALL PROGRAMS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. NO CONDITIONS, WARRANTIES OR REPRESENTATIONS SHALL BE DEEMED TO HAVE BEEN GIVEN BY VIRIDENT IN RESPECT OF THE PROGRAM OR DOCUMENTATION. VIRIDENT DOES NOT WARRANT OR ENTER INTO ANY TERM TO THE EFFECT THAT ANY PROGRAM WILL BE FREE FROM DEFECTS OR THAT ITS OPERATION WILL BE ERROR-FREE. ALL IMPLIED WARRANTIES, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR PURPOSE, NONINFRINGEMENT AND SATISFACTORY QUALITY ARE EXPRESSLY DISCLAIMED. If you have purchased technical support then errors will be addressed in accordance with the applicable technical support policies.

6. Limitation of Liability

VIRIDENT'S TOTAL AGGREGATE LIABILITY FOR ANY AND ALL CLAIMS SHALL NOT EXCEED THE AMOUNTS PAID FOR THE PROGRAMS DURING THE SIX-MONTH

PERIOD PRIOR TO THE DATE ON WHICH THE CLAIM AROSE. IN NO EVENT SHALL VIRIDENT BE LIABLE TO ANYONE FOR SPECIAL, COLLATERAL, INDIRECT, EXEMPLARY, INCIDENTAL OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING COSTS OF COVER, COSTS OF REWORK, RETESTING OR REMOVAL AND REINSTALLATION OF PRODUCTS, LOSS OF GOODWILL, LOSS OF PROFITS, OR LOSS OF USE) IN CONNECTION WITH THIS AGREEMENT ON ANY THEORY OF LIABILITY AND WHETHER OR NOT VIRIDENT HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES

7. Termination

Your license to use the Program(s) automatically terminates if You fail to comply with any provision of this agreement.

8. Assignment

You may assign all of Your rights to use the Program(a) to another person or entity provided that: (a) You transfer this agreement, the physical medium in which the Program is made available, user documentation and all copies of the Program(s) and user documentation together with the equipment on which it is installed; (b) You retain no copies of the Program(s), including copies stored on any medium; and (c) You remain liable under this Agreement for any breach of its terms by the assignee. Otherwise, the Programs may not be assigned without the consent of Virident.

9. Statutory rights

If You purchased the copy of the Program(s) covered by this Agreement as a consumer, the exclusions and limitations contained in this Agreement do not affect any statutory rights to which You are entitled and which may not by law be excluded or limited.

10. Certain Applications

Virident's Products are not designed and not warranted to be suitable for use in military applications and/or military environments, life support applications or other critical applications which may involve potential risk of death, personal injury or severe property or environmental damage Use of Products in such applications is fully at the risk of Buyer.

11. U.S. Government License Rights

If You are an entity of the United States Government you receive the Programs with restricted rights.. In no event shall the Government acquire greater than RESTRICTED/LIMITED RIGHTS. At a minimum, use, duplication, or disclosure by the Government is subject to restrictions as set forth in FAR §52.227-14 Alternates I, II and III (JUN 1987); FAR §52.227-19 (JUN 1987) and/or FAR §12.211/12.212 (Commercial Technical Data/Computer Software); and DFARS §252.227-7015 (NOV 1995) (Technical Data) and/or DFARS §227.7202 (Computer Software) as applicable. If Products are sold under a US Government contract, Virident rejects provisions or clauses required to be passed on to sellers pursuant to such contract and such provisions or clauses shall not be deemed to be binding on Virident unless accepted in writing by an authorized representative.

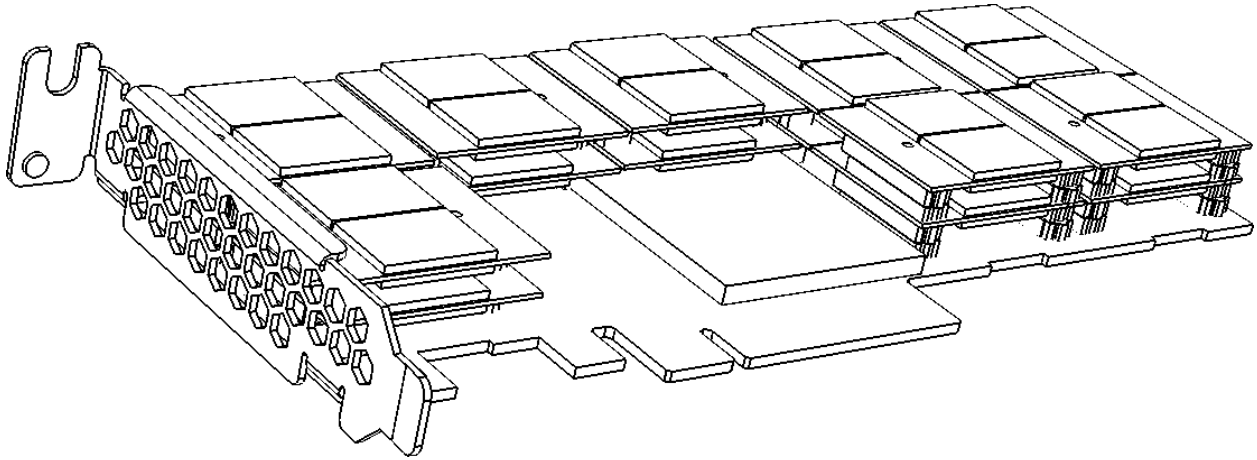
12. Miscellaneous

A. Export Control: Virident's obligation to provide the Products, documentation, or any media in which any of the foregoing is contained, as well as any other technical assistance shall be subject in all respects to such United States laws and regulations as shall

form time to time govern the license and delivery of technology and products abroad by persons subject to the jurisdiction of the United States, including the Export Administration Act of 1979, as amended, any successors or legislation, and the Export Administration Regulations and other United States laws and regulations governing exports in effect from time to time. **B. Governing Law and Jurisdiction:** This Agreement is deemed to have been entered into in the State of California, and its interpretation, construction, and the remedies for its enforcement or breach are to be applied pursuant to and in accordance with the laws of the State of California, excluding its conflict of law provisions. The UN Convention on the International Sale of Goods does not apply to this Agreement. The parties agree that in the event any action, suit or proceeding is brought by either party to enforce this Agreement, then such action, suit or proceeding may be brought only in a United States District Court for the Northern District of California, or, if that court does not have or elect to exercise its jurisdiction, then in the appropriate state court in the State of California, county of Santa Clara. **C. Contingency:** Virident is not liable, either wholly or in part, for nonperformance or a delay in performance due to force majeure or contingencies or causes beyond the reasonable control of Virident, including but not

limited to, shortage of labor, fuel, raw material or machinery or technical or yield failure where Virident has exercised ordinary care in the prevention thereof. Production and deliveries may be allocated by Virident in a reasonable manner in the event of shortage of goods. **D. Partial Invalidity:** If any provision of this Agreement is held to be invalid or unenforceable, the remainder of this Agreement shall continue in full force and effect so long as neither party is denied material economic benefit hereunder and will be interpreted to reflect the original intent of the parties. **E. Waiver.** A waiver by Virident of any term or condition of this Agreement in any instance shall not be deemed or construed to be a future waiver of such term or condition. **G. Construction; Language.** As used in this Agreement, the words "include" and "including," and variations thereof, will not be deemed to be terms of limitation, but rather will be deemed to be followed by the words "without limitation." The headings in this Agreement will not be referred to in connection with the construction or interpretation of this Agreement. This Agreement is in the English language only, which language shall be controlling in all respects, and all notices under this Agreement shall be in the English language.

Introduction



The Virident tachION Drive is a revolutionary solid-state storage product utilizes an industry standard PCI Express interface and an innovative hardware-software architecture for high performance sustained IOPS. TachION delivers best-in-class performance and enterprise reliability, low power together with plug-and-play ease-of-use for performance starved, data-intensive workloads such as databases, business analytics, simulation, visualization and high-performance computing.

The tachION drive comes in capacities ranging from 200GB up to 400GB in a single, half-height half-length form factor. This capacity is available to applications at bandwidths exceeding 1.3 GB/s for reads and over 800 MB/s for standard 75:25 read:write mixes at 4K block sizes under full capacity and full load.

This document explains how to install, configure, benchmark, and manage the tachION product for Linux systems.

Specifications

Virident tachION Specifications		
	Capacity (GB)	200, 300, 400
	NAND Flash	SLC
	Bandwidth	1.3 GB/s (Peak)
	Read Performance	1.3 GB/s (4KB blocks) 320,000 IOPS
	Peak Read/Write Performance (75:25 read/write)	1.3 GB/s (4KB blocks) 320,000 IOPS
	Sustained Read/Write Performance (75:25 read/write)	800 MB/s (sustained) 200,000 IOPS
	Advanced ECC	End-to-end error correction Global and local wear-leveling BCH (16 bits / 256 bytes) RAID 0, 1, 5, 10
	Lifetime	24 years (@ 5 TB writes/day)
	Enterprise Class Reliability	UBER < 10 ⁻¹⁷
	Operating Temperature	0°C/32°F to 50°C/122°F
	Non-operating Temperature	-40°C/-40°F to 70°C/158°F
	Airflow (LFM)	200
	Humidity	5% to 95% (relative, non-condensing)

System Requirements

In order to utilize the Virident tachION card your system must meet the following requirements.

Hardware

- PCI-Express x8/x16 electrical slot, half- or full-height
- Minimum of 200 LFM airflow over card
- 2- or 4-socket, multicore X86 server processor
- 4GB of memory minimum

For optimal performance we recommend a system utilizing the Intel™ Xeon 5500 CPU and X58 chipsets.

Operating System Support

- Red Hat Enterprise Linux (RHEL) 5.2 or later
- CentOS 5.2 or later
- SUSE Linux Enterprise Server (SLES) 10SP3, 11 or later

Installation

Before beginning please unpack the product box and ensure you have the following:

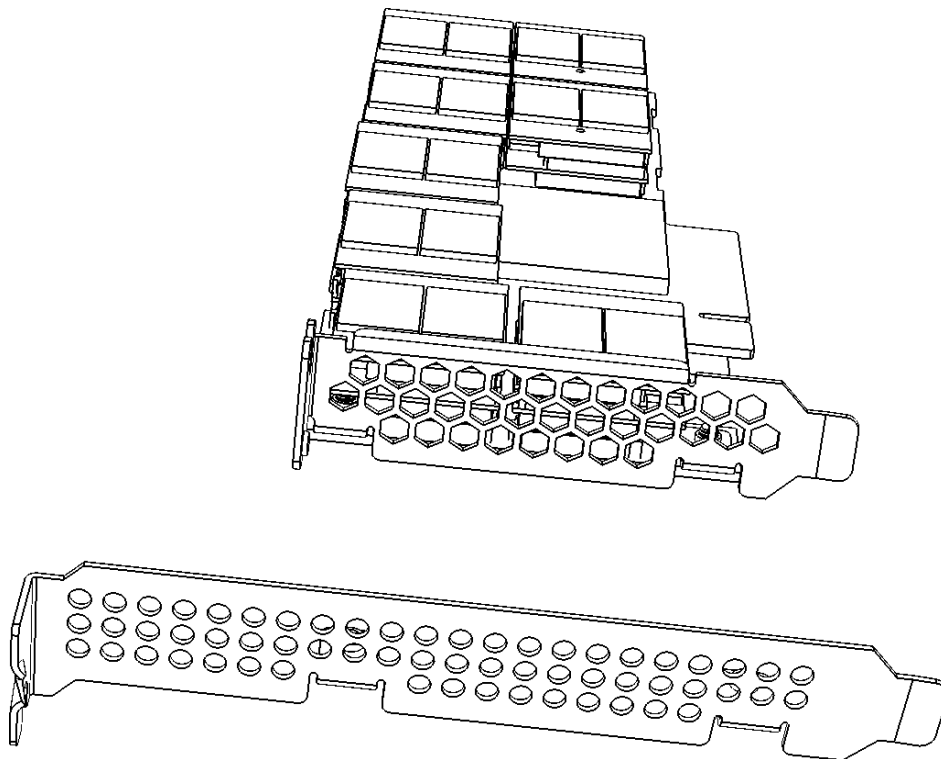
- Virident tachION drive card at 200, 300, or 400GB capacity
- Half-height and full-height PCI bracket
- USB driver stick
- User's Guide

Hardware installation

Note: Please observe all standard ESD precautions while installing the Virident tachION card in the server. Electrostatic discharge can damage the components used in the card and the server itself.

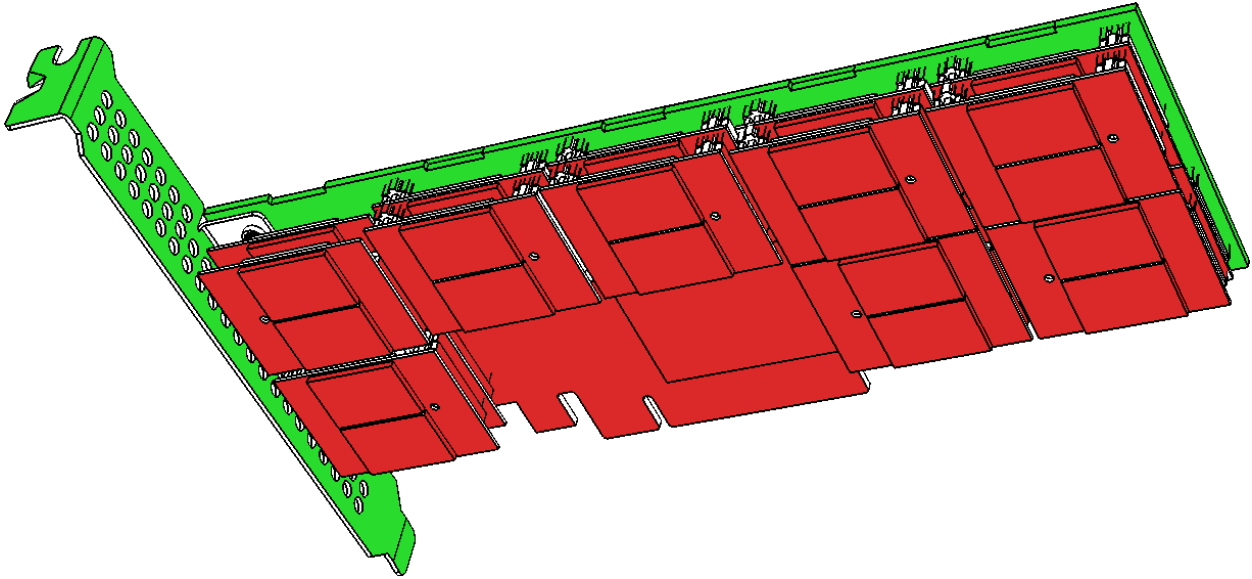
First, as for all upgrades, please power down the system following your operating system's standard recommendations. Unplug the server from the AC power supply to ensure no dangerous voltages are present.

Install the proper bracket on the Virident tachION card. The card ships from the factory with a low-profile bracket already installed. If you have a full-height PCI express slot please install the included full-height bracket before installation by removing the mounting screws from the back of the card, replacing the bracket, and re-tightening the screws evenly.





Please do not exert force or grip the daughter boards (highlighted in red below). Hold the card only by the PCI bracket and the top and bottom edges of the PCB, highlighted in green below. If force is applied to the daughter boards they may disconnect from the main Virident tachIOn PCB, or even have their connectors damaged.



Select a free slot on the motherboard to install the card. Now is a good time to verify the slot size is x8 electrically. Often the slot electrical width is stenciled in white next to each connector. A x16 physical length slot is also supported but will provide no additional performance.



IMPORTANT: Some motherboards have slots which are physically x8 in length but only x4 electrically. Installing the card in such a slot will result in unpredictable or significantly lower performance.

Software installation

The Virident tachIOn comes with a standard, FAT formatted USB stick containing the necessary device drivers and test scripts to ensure proper operation of the device. For driver updates please check the Virident support site, <http://support.virident.com>.

All the following installation instructions need to be executed as the root user. Please login as root or “sudo bash” as appropriate before beginning. There will also be a reboot required to enable certain kernel parameter changes so please set aside fifteen minutes of scheduled downtime for these operations.

Verify the Virident tachIOn is installed and detected properly

After powering on the system for the first time with the card installed, please verify that it is detected and initialized properly by the PCI subsystem by using the “lspci” command. If the card is not detected please power down the server and ensure it is seated properly. If it is still not detected please contact Virident support for further assistance.

```
[root@hw01 ~]# /sbin/lspci -d 1a78:
04:00.0 FLASH memory: Virident Systems Inc. Unknown device 0031 (rev 01)
```

Choosing and installing the proper driver RPMs

The USB driver stick will contain multiple versions of the Virident tachIOn drivers for different kernels and operating systems. Before installing one, please verify your running kernel version with the “uname” command:

```
[root@hw01 ~]# uname -r
2.6.18-164.el5
```

In the mounted driver stick there will be individual directories, each with several RPMs. Select the directory corresponding to the running kernel and install all RPMs within it. There will be a “vgc-firmware-<revision>.rpm”, “vgc-utils-<revision>.rpm”, and “vgc-drivers-<kernel-version>-<revision>.rpm”. Install these using the standard RPM commands:

```
[root@hw01 RPMS]# rpm -ivh \
    /mnt/usb/2.6.18-164.el5/vgc-drivers-2.6.18-164.el5-1.0.x86_64.rpm \
    /mnt/usb/2.6.18-164.el5/vgc-utils-1.0.x86_64.rpm \
    /mnt/usb/2.6.18-164.el5/vgc-firmware-1.0.noarch.rpm

Preparing...                               ##### [100%]
 1:vgc-firmware                             ##### [ 33%]
 2:vgc-drivers-2.6.18-164.##### [67%]
* INFO: service irqbalance disabled *
 3:vgc-utils                               ##### [100%]
```

If you do not find a driver version corresponding to your current kernel, please contact Virident support for more information about compiling from sources. Forcing installation of a device driver from a different kernel version can often result in improper behavior and system instability.

SUSE Only: Enable unsupported modules to allow driver loading



While the Virident tachIO device driver is fully tested under SUSE Linux, it is not supported by Novell. By default SUSE will not allow unsupported modules to be loaded into the kernel, so this setting must be changed to allow the Virident device driver to load. Edit the file `"/etc/modprobe.d/unsupported-modules"` and set the value `"allow_unsupported_modules"` to `"1"` as shown below. Without this change, the device driver will be unable to load.

```
[root@hw01 RPMS]# vi /etc/modprobe.d/unsupported-modules
...
# Setting allow_unsupported_modules 1 enables loading of unsupported
# by modprobe, setting allow_unsupported_modules 0 disables it. This can
# be overridden using the --allow-unsupported-modules commandline switch.

allow_unsupported_modules 1
```

RECOMMENDED: Enhance performance by disabling CPU throttling

To obtain maximum performance CPU throttling should be disabled by the operating system. This is especially important for benchmarking and high sustained IO workloads. The CPU throttling agent can be disabled by the following operation:

```
[root@hw01 ~]# chkconfig cpuspeed off
[root@hw01 ~]# chkconfig --list cpuspeed
cpuspeed      0:off  1:on   2:off  3:off  4:off  5:off  6:off
```

Update kernel options and reboot

While the Virident tachION device driver can run without any special kernel options, to get the absolute highest performance from the product we require the setting of two kernel boot-time options, “idle=poll” and “isolcpus=3”.

Please add these parameters to the GRUB configuration file “/boot/grub/menu.lst” as shown below:

```
[root@hw01 RPMS]# vi /boot/grub/menu.lst
...
title Red Hat Enterprise Linux Server (2.6.18-164.el5)
    root (hd0,0)
    kernel /vmlinuz-2.6.18-164 ro root=/dev/sda3 isolcpus=3 idle=poll
    initrd /initrd-2.6.18-164.img
```

These parameters reduce the interrupt latency and significantly improve the current device driver performance. Further information about their purpose is available in your operating system’s kernel documentation folder.

Once the configuration file has been edited reboot the system and enter your BIOS setup utility for further optimization.

BIOS setup for optimal performance

When your system restarts, and before booting the operating system, please enter its system BIOS menus (often done by pressing the Delete or F2 keys) to verify several important settings. Please note that each BIOS often has its own nomenclature for these settings, so the exact wording of the configuration parameters may not match the examples. Not all BIOSes will have all of these options, especially on AMD based systems, so if a parameter isn’t available simply skip to the next one as the Virident tachION card will still function properly, at slightly less than full performance.

If there are any concerns please contact Virident support which will be able to walk you through the operations.

The following settings are often found under the “Advanced : Processor/Chipset Control” BIOS menus.

Disable the following optional settings:

- Processor Performance States / C States / C-State Tech / EIST
- C1E States / C1 Enhanced States
- (Link) Active State Power Management
- DRAM node interleaving

Enable the following on Intel™ based systems:

- Hyperthreading / Logical Processors
- Turbo Mode

Finally, if there are performance profiles (System Power Modes) available set them to “Maximum Performance.”

Once these settings are verified save the changes and let the system fully boot.

Verify the device driver is loaded and functioning

If messages are enabled during your boot there will be a line in the boot log similar to the following:

```
Loading tachION kernel modules...          [ OK ]
Starting tachION monitoring service...       [ OK ]
```

Log in again and become root and verify that the new Virident tachION device is present using “ls” or “fdisk -l”:

```
[root@hw01 RPMS]# ls -l /dev/vgca0
brw-r----- 1 root disk 252, 0 May 27 09:45 /dev/vgca0

[root@hw01 RPMS]# fdisk -l /dev/vgca0

Disk /dev/vgca0: 401.3 GB, 401327783936 bytes
186 heads, 128 sectors/track, 32923 cylinders
Units = cylinders of 23808 * 512 = 12189696 bytes

Disk /dev/vgca0 doesn't contain a valid partition table
```

If the device is not present please refer to the troubleshooting section of this guide for more assistance.

Verify installation performance

Before making a filesystem on the Virident tachIon drive, it is useful to benchmark the raw performance of the entire system to ensure proper setup. That way, if any performance problems appear they can be isolated and ameliorated more easily.

Included on the USB driver stick is a simple script which utilizes the GPL licensed FIO IO benchmark. This test script verifies things from RPM installation, to PCI express link lanes, to absolute read and write IOPS. Copy the required files “test.sh” and “fio” to the current directory and run it as shown below. Should there be any noted, significant performance problems please check the Troubleshooting and Performance Optimization sections of this guide, or contact Virident support.

```
[root@hw01 ~]# cp /mnt/usb/test.sh /mnt/usb/fio .
[root@hw01 ~]# bash ./test.sh
*** VIRIDENT TACHION PERFORMANCE TEST ***

-----
WARNING! WARNING! WARNING! WARNING! WARNING! WARNING! WARNING! WARNING!
THIS TEST WILL DESTROY ANY DATA AND FILESYSTEMS ON THE VIRIDENT DEVICE.
Type the word yes and hit return to continue, or anything else to abort.
-----

Really run performance test? [no/yes]  yes
Card is plugged into x8 slot...OK
Checking boot option isolcpus=3  for high performance...OK
Checking idle=poll for high performance...OK
Unmounting filesystem on /dev/vgca0...OK
-----

Running Max Write Bandwidth Test....
Max Write Bandwidth measured is 1160 MiB/s while expected 900 MiB/s...OK
-----

Running Max Read Bandwidth Test....
Max Read Bandwidth measured is 1330 MiB/s while expected 1200 MiB/s....OK
-----

Running Max 4K Read IOPS Test....
Max Read IOPS measured is 332 K while expected average is 300 K ...OK
-----

Running Max 4K Write IOPS Test....
Max Write IOPS measured is 149 K while expected average is 100 K ...OK
-----
```

Use the “/dev/vgca0” device

Congratulations, the Virident tachIon is now ready to be utilized! Please read the following sections for partitioning and filesystem creation options to ensure optimal performance.

Partitioning and Configuring with VGC-Config

NOTE: Most use cases will not need to partition the device or change the performance balance between capacity and write bandwidth. It is very often sufficient to create a filesystem directly on the “/dev/vgca0” device. To do so, please refer to the following Filesystems section.



IMPORTANT: In the cases when it is necessary to partition the Virident tachION drive, please do not use “fdisk” as it is not flash aware and can result in significant performance problems. Fdisk is also not be able to support additional features like adjusting the write balance of the card, nor will it be capable of enabling future on-board hardware RAID on the Virident tachION drive.

Partitioning and selecting write performance of the device

One, two, or four partitions may be created on the Virident tachION card using the vgc-config utility. Each of these partitions can have different capacity/write performance characteristics as well, so one partition could be optimized to have maximum space available for static user data while another can be optimized for heavily updated data (like database logs, scratch space, etc.).

Usable space vs. write performance tradeoffs

By default the Virident tachION drive creates partitions with a “balanced” setting that is optimal for most real-world workloads. A small portion of the raw capacity of each tachION daughterboard is set aside to help accelerate writes, while the majority is used to store user data.

To fine-tune application performance Virident allows customizing this ratio of usable data to write acceleration area. By increasing the amount set aside for writes in “maxperformance” mode, the Virident tachION drive is able to sustain an even higher write workload at the cost of reduced usable capacity. Conversely, by decreasing the amount set aside for write acceleration in “maxcapacity” mode it is possible to allow more user data to be stored on the device but with lower sustained write performance.

Performance implications

Please note that the act of partitioning physically segregates the data to individual tachION daughter boards. While this is useful for data protection it does reduce the attainable performance for each individual partition to ½ or ¼ of the total card performance.

VGC-Config example

Below is the usage information for vgc-config. While additional options are available within the utility, most users will only need to select a number of partitions (“-n XX”) and a write performance profile (“-m YYY”) to set up the Virident tachION card for use.

```
[root@hw01 ~]# vgc-config -h

Usage: vgc-config -d <device name> [ [-l]|[-nms]|[-r]|[-i]|[-u] ] or
      vgc-config -p <partition name> [ [-l]|[-ms]|[-r]|[-i]|[-u] ]
      -d : device name /dev/vgc[X]
      -p : partition name /dev/vgc[X][Y]
      -l : list current and pending config changes
      -n : number of partitions, (1/2/4)
      -m : balanced/maxperformance/maxcapacity mode
      -s : block device sector size, 512(default)/4096
      -r : reset to factory defaults
      -i : interactive mode
      -u : undo pending config changes
      -h : display this help and exit
```

The following example show how to create two partitions with the same performance profile and reload the configuration file to ensure the change takes effect.


```

[root@hw01 RPMS]# vgc-config -d /dev/vgca0 -n 2 -m maxperformance

Reload 'vgcd' service to apply the new configuration changes.
[cmd: service vgcd reload]
Warning: Applying new configuration changes may erase existing data on
'/dev/vgca'. Please backup the data on '/dev/vgca'
Note: Use vgc-config '-u' option to undo the changes.

[root@hw01 RPMS]# service vgcd reload
Stopping TachION monitoring service...          [ OK ]
Unloading kernel modules...                     [ OK ]
Committing new configuration                    [ START ]
INFO: Configuring TachION card a ...
INFO: Changing number of partitions on TachION card a to 2
Executing vgcsave for device a
vgcsave successfully completed on device a. All previously stored data
on formatted device may be lost.
INFO: Preparing device /dev/vgca0 with options -g 33
INFO: Prepared /dev/vgca0 successfully
INFO: Preparing device /dev/vgca1 with options -g 33
INFO: Prepared /dev/vgca1 successfully
INFO: Device configuration completed successfully
Committed new configuration                      [ OK ]
INFO: kernel modules are already loaded
Starting TachION monitoring service...           [ OK ]
For more information about the additional options please see the appendix
and man pages for the vgc-config utility.

[root@hw01 RPMS]# vgc-config -d /dev/vgca0 -l
Current Configuration:
/dev/vgca      2 partitions
    /dev/vgca0      mode=maxperformance sector-size=512
    /dev/vgca1      mode=maxperformance sector-size=512

```

The first command, “vgc-config -d /dev/vgca0 -n 2 -m maxperformance,” only sets up a partitioning scheme but does not implement it (as this will cause data on the card to be erased).

The second command, “service vgcd reload,” actually performs the partitioning and allocation, erasing any existing data on the card. Repartitioning can be undone at any time before a reboot or “service vgcd reload” by simply executing “vgc-config -u”.

For more detailed examples of vgc-config please see the man pages included with the driver release.

Creating Filesystems

Any standard Linux filesystem may be placed on the Virident tachIO card simply by using its standard mkfs utility. Virident recommends using the XFS filesystem if possible as it can deliver more of the raw throughput of the Virident tachIO card with less tweaking, but EXT3 or other filesystems can also give excellent results if care is taken

This section goes into filesystem-specific options that will improve the performance of the card for most applications.

EXT3 high performance options

At high write workloads EXT3 can bottleneck on writes to its journal, especially on devices as high in performance as the Virident tachIO drive. To help eliminate this bottleneck Virident recommends that additional space be set aside for the EXT3 journal using the following option

```
mkfs.ext3 -J size=400 <other standard options as required> /dev/vgca0
```

Another alternative, applicable in cases where the EXT3 journal is not needed at all such as scratch-areas, is to simply use EXT2 which has no such journaling.

XFS high performance options

XFS addresses some of the performance limiting bottlenecks of EXT3, and is available for most kernels. It offers significantly increased multithreaded, single-file write bandwidth which often accelerates database performance in particular.

Virident recommends the following standard mkfs command line for XFS to enable the highest write performance:

```
mkfs.xfs -s size=4096 <other standard options as required> /dev/vgca0
```

Software RAID

Careful partitioning, software RAID creation, and filesystem options can provide very good performance on the Virident tachION card. This section provides several recipes for known performing configurations.

RAID-1

With RAID-1 (mirroring), the default RAID creation parameters are sufficient for good read and write performance. Please ensure you are utilizing the proper filesystem options (specified in the “Creating Filesystems” section of this document) to ensure maximum performance.

```
mdadm --create /dev/md0 --level=1 --raid-devices=2 /dev/vgca0 /dev/vgca1
```

RAID-5/6

RAID-5/6 write performance is very critical, since data is striped across cards with parity and modification of a single byte will result in a corresponding parity update write. XFS has been shown to have much better performance than EXT3 in this mode.

The default chunk size for RAID-5/6 is unnecessarily large for the Virident tachION drive at 64KB. Such a large chunk size often wastes bandwidth on read-modify-writes for normal application-sized writes. Instead, for the Virident tachION drive a chunk size of 4KB is most optimal as this will limit the amount of wasted bandwidth and more fully utilize the high random read-write capabilities of the device.

The filesystem also needs to be made aware of this RAID striping to allow it to minimize read-modify-writes by sending data to the block devices in (chunk*stripe)-sized units. XFS has the parameters “su” and “sw” to set the stripe unit and stripe width, respectively.

The following command lines illustrate the above settings

```
# Create RAID-5
mdadm --create /dev/md0 --level=5 --raid-devices=4 --chunk=4 \
        /dev/vgca0 /dev/vgca1 /dev/vgca2 /dev/vgca3

# Create RAID-6
mdadm --create /dev/md0 --level=6 --raid-devices=4 --chunk=4 \
        /dev/vgca0 /dev/vgca1 /dev/vgca2 /dev/vgca3

# Make XFS with proper RAID awareness
mkfs.xfs -f -d su=4096 -d sw=8192 -d agcount=32 /dev/md0
```

Multiple Virident tachIOon Cards

Up to eight cards are supported in a single system and are automatically supported by the standard device driver. Each device will appear as `"/dev/vgc[a..h]0"` and can be accessed individually or combined using software RAID.

The kernel boot option `"isolcpus=3,"` added during the card install stages, needs to be modified to provide optimal read and write performance when utilizing two or more cards. Please contact Virident support for assistance in optimizing these configurations.

TachION Management Utilities

The state of each card is monitored on an ongoing basis by the device driver and logged to stable storage. This accurate and complete history helps identify any issues before they become real problems. A suite of utilities is delivered with the device driver to allow end-users access to this performance history.

VGC-Config

Most administrative interactions with the Virident tachION card are handled by the “vgc-config” utility. This utility can show the current partitioning and write acceleration modes on an existing card.

To change settings for the entire Virident tachION card use the “-d /dev/vgca” option. To modify the state of single partitions use “-p /dev/vgca[0..3]”. Partitions are numbered starting at 0 and incrementing after the card letter (normally “a” when only a single card is installed).

All state-changing operations require the execution of “service vgcd reload” to take effect. This allows a user to ensure that any required data is backed up off of the Virident tachION drive before when repartitioning takes place. Once repartitions are committed with the “reload” operation all data on the card will be erased unrecoverably.

For multiple card installations this “service vgcd reload” command will cause the device driver to be reloaded as part of the partitioning process. Data on unmodified cards will not be lost, however it is necessary to unmount any filesystems present on unmodified cards to allow the device driver to reload.

VGC-Config card settings

The entire card is modified when “-d /dev/vgca” is specified, and the following options may be used. Note that changes will not occur until a “service vgcd reload” is performed.

VCG-CONFIG Per-Card Settings	
-l	Lists the current partition and write acceleration modes, as well as any uncommitted partition or acceleration mode changes
-r	Resets the card to factory default settings (single partition, balanced write mode)
-u	Undoes any uncommitted partition or acceleration mode changes. Use this option when an inadvertent partitioning change has been made, but before any “service vgcd reload” has been performed. Once the service has been reloaded any partition changes are no longer undoable.
-n #	Sets the number of partitions to create on the card. Only powers-of-two may be allowed, and only up to 4 partitions are supported presently. All partitions are created with equal size.
-m MODE	Specifies the write acceleration mode (either balanced, maxperformance, or maxcapacity) for all partitions on the device
-s #	Sets the hardware sector size for all partitions on the card. Most software and filesystems expect 512 bytes, so do not change this setting unless you are sure the application can handle the larger, 4K sectors.

VGC-Config partition settings

Individual partitions may be customized using “-p /dev/vgca[0..3]” to change things like sector size or write acceleration mode. Only the following options may be specified in this mode:

VCG-CONFIG Per-Partition Settings	
-m MODE	Specifies the write acceleration mode (either balanced, maxperformance, or maxcapacity) for the specified partition
-s #	Sets the hardware sector size for the specified partition. Most software and filesystems expect 512 bytes, so do not change this setting unless you are sure the application can handle the larger, 4K sectors.

VGC-Diags

A simple to use diagnostic reporting utility, “vgc-diags” is included with the device driver. Virident support may ask you to run it in “-vv” very verbose mode when helping bring up the system, but it may be used at any time to get some general card information (not related to usage).

```
[root@hw01 virident]# vgc-diags -d /dev/vgca0
Checking Software configuration:
-----
Found: vgc-firmware-1.0 vgc-utils-1.0 vgc-drivers-2.6.18-164.3.1.el5
Checking Hardware configuration:
-----
Virident TachION a,      TachION0: 0x8000000000 bytes (512 GB)
BAR [0] phys 0xdf100000 virt 0xfffffc20010180000 (sz 1 MB)
BAR [1] phys 0xd4000000 (sz 1 KB)
      Card Name                      = TachION LP
      TachION Controller RTL rev      = 21812
      TachION Controller ucode rev    = 21328
      TachION Controller frequency    = 70 MHz
      Module Controller RTL rev       = 21708
      Module Controller ucode rev     = 21708
      Card Temperature                = 62 C
      CPU Frequency                   = 2261 MHz
```

Performance Optimization

Once the card driver is installed and running there are certain generic optimizations that may be made to the system which can deliver improved results.

Disabling power saving and enabling performance CPU modes

Disabling power saving modes on the CPU can often deliver vastly improved performance on systems with very high IO loads. The clock throttling can be slow to respond to increased CPU demands and can interfere with the speed at which interrupts are processed which can have a very high impact on IO latencies.

CPU throttling is controlled daemon which may be disabled via:

```
chkconfig cpuspeed off
service cpuspeed stop
```

Enabling high performance modes on the CPU is often done automatically by the operating system once the CPU throttling daemon is disabled. To explicitly request the highest-performance modes you may execute the following command (change the 15 to the number of total virtual/physical cores you have on the system)

```
for cpu in `seq 0 15`; do
    cpufreq-selector -c $cpu -g performance
done
```

Increasing thread counts to boost performance

In many applications the number of threads dedicated to IO is configurable. Most configurations, however, assume that IO is relatively slow and therefore doesn't need many CPU resources dedicated to it. Since the Virident tachION drive can perform many orders of magnitude higher than standard storage arrays, this assumption is no longer correct. Linearly increasing the number of IO threads running against a tachION card can often linearly increase the IO bandwidth delivered to the application.

MySQL specific optimization

Both the Google and Percona patched versions of MySQL InnoDB and XtraDB support increasing IO threads. Through extensive testing Virident recommends adding the following setting in /etc/my.cnf to enable the highest performance on the Virident tachION drive. Please contact Virident support for more information on optimizing MySQL on the Virident tachION card.

```
/etc/my.cnf
...
[mysqld]
innodb_write_io_threads=8
innodb_read_io_threads=8
innodb_io_capacity=10000
```

Filesystem selection

The filesystem chosen to store data can have a dramatic impact on application performance. Some filesystems have been designed and optimized for slower, lower performing storage technologies. Unfortunately many of those optimizations for slower speed devices can reduce the maximum performance available on a high performance storage subsystem.

As prerequisite, always create the filesystems with the suggested options (see the Creating Filesystems). For filesystems not listed it is recommended to enable any options which increase the minimum sector size from 512 bytes to 4096 bytes if possible. Also, of course, mount with “-oatime” to disable file metadata updates on each access.

EXT3 has a performance issue when multiple threads are writing to a single file. This is due to an internal locking mechanism which only allows a single thread to write to a file at any one time, even if those two threads are trying to write to completely different sections of the file. As a result, the maximum achievable write performance is much lower for a single file than the underlying raw is capable. Each individual file has its own lock, so if many smaller files are being accesses in parallel this bottleneck is not nearly as evident. Reads on EXT3 are unaffected by this lock and do proceed at device speed, even for multiple readers.

Virident suggests using XFS on the Virident tachION drive if possible. XFS has been seen to deliver much higher single-file write performance because the EXT3 write bottleneck has been effectively broken in this filesystem, so a single file can deliver the same write performance as multiple files. Other optimizations have shown reduced filesystem overhead, even for reads.

Troubleshooting

Cannot find block device /dev/vgca0

Ensure the card is detected on the PCI bus. If it is not please power down the system and check the card seating in the slot. If it is still undetected please contact Virident support for replacement procedures.

```
[root@hw01 ~]# /sbin/lspci -d 1a78:
04:00.0 FLASH memory: Virident Systems Inc. Unknown device 0031 (rev 01)
```

Ensure the device driver RPMs are installed and available. If the running kernel version and the device driver kernel version do not match then the device driver will be unable to load. Verify the “uname -r” output is the same as the installed RPM version and the driver files are visible and loadable. The manual “modprobe vgcinit” command should return without any errors.

```
[root@hw01 ~]# uname -r
2.6.18-168.el5
[root@hw01 ~]# ls -l /lib/modules/`uname -r`/vgc
total 14180
-rw-r--r--. 1 root root 410672 May 28 14:16 vgcdebug.ko
-rw-r--r--. 1 root root 4789032 May 28 14:16 vgcdrive.ko
-rw-r--r--. 1 root root 312640 May 28 14:16 vgcinit.ko
-rw-r--r--. 1 root root 241040 May 28 14:16 vgcmgmt.ko
-rw-r--r--. 1 root root 509600 May 28 14:15 vgcmttd.ko
-rw-r--r--. 1 root root 5106736 May 28 14:16 vgcport.ko
-rw-r--r--. 1 root root 3133920 May 28 14:16 vgcsch.ko
[root@hw01 ~]# modprobe vgcinit
[root@hw01 ~]#
```

Ensure the service is started. If not, then manually start it with “service vgcd start”.

```
[root@hw01 ~]# service vgcd status
INFO: kernel modules are loaded
```

Ensure the card is configured. If not, please manually configure it with a single partition (see the Partitioning and Configuration section for more details) and try once again.

```
[root@hw01 ~]# vgc-config
Current Configuration:
/dev/vgca          1 partition
                   /dev/vgca0          mode=balanced          sector-size=512
```

Performance is lower than expected

If the initial test script reports lower than expected performance, or your own application is seeing lower than expected IOPs, please begin by verifying the card is installed in a true PCI-Express x8 electrical slot. Certain motherboards provide x8 slots which are physically the right size but only connect 1 or 4 of the required PCI express lanes. Your motherboard manual or the motherboard itself is the best reference for this. Often the electrical connections are listed in white lettering next to the slot on motherboards, if the manual is not handy.

Next please verify the BIOS and kernel settings requested in the Installation section. Of particular importance are the disabling of the power savings modes and enabling of performance modes in the BIOS. Check the `/proc/cmdline` file to get the current running kernel's options. If the "isolcpus" and "idle" options are not shown in this file, then either the system has not been rebooted or the incorrect `/boot/grub/menu.lst` configuration entry was modified.

Finally, filesystems can have a large impact on performance. Review the "Creating Filesystems" section for more information on the recommended options for XFS and EXT3.

IRQBalance not running

In order to provide maximum performance, the Virident tachION card requires a fixed core to handle interrupt processing. To guarantee this allocation it is necessary to disable the IRQ balancing service. The IRQ balancer is re-enabled when the device driver RPMs are removed.

Appendix A: Onboard diagnostic LEDs

There are three onboard LEDs on the Virident tachION card. They are visible through the back bracket and always indicate the state of the system. These LEDs can also be used to identify individual cards in a multi-card installation. See the Multiple Virident tachION Cards section for more details on this use.

The following table explains what each combination of these LEDs indicates:

Green	Yellow	Amber	Status Indicated
OFF	OFF	OFF	No power to the card.
ON	OFF	ON	Power on, no driver loaded.
ON	OFF	OFF	Power on, driver loaded.
ON	FLASH	OFF	Data being written.
FLASH	OFF	OFF	Data being read.
ON	FLASH	ON	Card firmware being updated. Do not remove power from the system.

Contacting Support

- Email: support@virident.com
- Phone: 1-866-395-9606